Supplemental Material. Calculation of the *modified* – Benchmark Dose

The general form of the models referred to in the Mathematical Modelling section is:

$$P(d) = c + (1-c)P^*(d) \quad [1]$$

where d is dose, P(d) is the response to dose, c is background at zero dose and $P^*(d)$ is a particular parametric model used to fit the data. The seven $P^*(d)$ are:

Weibull	$P^*(d) = 1-\exp(-(d/q_1)^{q_2})$
Gamma	$P^*(d) = \frac{1}{\Gamma(q_2)}\gamma(q_2, d/q_1)$
Log Normal	$P^*(d) = \frac{1}{2}[1+erf(\log(d/q_1)/q_2\sqrt{2})$
Log Logistic	$P^*(d) = (d/q_1)^{q_2}/(1+(d/q_1)^{q_2})$
Truncated Normal	$P^*(d) = [P^{**}(d)-P^{**}(0)]/[1-P^{**}(0)]$
	$P^{**}(d) = \frac{1}{2}[1+erf((d-q_2)/q_1\sqrt{2})]$
Truncated Logistic	$P^*(d) = [P^{**}(d)-P^{**}(0)]/[1-P^{**}(0)]$
	$P^{**}(d) = 1/[1+\exp(-(d-q_2)/q_1)]$
Linear Exponential	$P^*(d) = 1-\exp(-(d/q_1))$ [2]

q_1 and q_2 are the two unknown parameters which define $P^*(d)$ and combine with c to give up to three unknowns when applicable. Here $erf(.)$ is the error function, $\gamma(.,.)$ is the incomplete gamma function and $\Gamma(.)$ is the gamma function.

The benchmark risk is the extra risk, $R(P(d))$, at dose d:

$$R(P(d)) = \frac{P(d)-P(0)}{1-P(0)} \quad [3]$$

$$= P^*(d) \quad [4]$$

The benchmark dose corresponding to a given R, in this case 0.05, is

$$d = (P^*)^{-1}R \quad [5]$$

The MLE is used to determine the parameters c, q_1, q_2 for each model based on the binomial likelihood function,

$$M = \prod_{i=1}^{N} {}^{n_i}C_{r_i} P_i^{r_i}(1-P_i)^{n_i-r_i} \quad [6]$$

and reduced log likelihood function

$$L = \sum_{i=1}^{N}\{r_i \log_e P_i + (n_i - r_i)\log_e(1-P_i)\} \quad [7]$$

with the omission of $\log_e({}^{n_i}C_{r_i})$. In [7], $P_i = P_i(d_i)$, with d_i the i^{th} dose of N doses and r_i is the number of responses as tumors in the group of n_i animals.

At a maximum of L, the gradient conditions are the non-linear equations

$$\frac{\partial L}{\partial \beta_k} = \sum_{i=1}^{N} n_i \left[\frac{y_i}{P_i} - \frac{1-y_i}{1-P_i}\right]\frac{\partial P_i}{\partial \beta_k} = 0 \ , \quad k = 1,2,3 \quad [8]$$

where the parameters c, q_1, q_2 are replaced by $\beta_1, \beta_2, \beta_3$ respectively and $y_i = r_i/n_i$.

The results of fitted parameters for the BaP data of Table 1 are given in Supplemental Material, Table 1.

In a first procedure, Newton Raphson iteration was used to determine $\beta_1, \beta_2, \beta_3$ from equation [8] . This requires second derivatives $\partial^2 L/\partial\beta_k\partial\beta_j$ and starting values for the parameters. Many different starting values were selected for the parameters in each model to check for the presence of more than one maximum in each model. However, only one maximum was found in each case. In a second procedure, values of MLE parameters were confirmed using a robust implementation of the Nelder-Mead simplex method (Miller, 2004). As referred to in the body of the text, the Truncated Normal could not be fitted at all, being near supralinearity. An indication of supralinearity is when the Weibull parameter q_2 is < 1, which is nearly so here.

To determine the $BMD_{0.05}$, the NHMRC (1999) guidelines recommend the use of a minimum of three models: Weibull, Linear Exponential, and either Log Normal or Log Logistic, or an extended suite of models as in [2]. For each model it is required that the goodness-of-fit to the data be assessed, both graphically and algebraically. However, there are no prescriptions given for the precision of such fits. After assessments, the $BMD_{0.05}$ is found by arithmetically averaging the $d_{0.05}$ values for well fitting models.

From the display of curves in Figure 1A, it appears that all the curves are similar and constitute good fits when compared with fits for other carcinogen dose response data.

To assess the fits algebraically, some standard measures (eg. Read and Cressie, 1988) are now listed: The Pearson chi-squared statistic

$$X^2 = \sum_{i=1}^{N} \frac{n_i(y_i - P_i)}{P_i(1-P_i)} \qquad [9]$$

and the log likelihood ratio

$$G^2 = 2\sum_{i=1}^{M} \left\{ n_i \log_e \frac{r_i}{n_i P_i} + (n_i - r_i)\log_e \frac{(n_i - r_i)}{n_i(1-P_i)} \right\} \qquad [10]$$

The p-value determined for each of these measures based on the chi-squared distribution, χ^2, is

$$p(\chi^2 \geq x) = 1 - \int_0^x u^{\nu/2} e^{-u/2} du \Big/ \left[2^{\nu/2}\Gamma(\nu/2) \right] \qquad [11]$$

Here, x is either value of X^2 or G^2 and ν is the number of degrees of freedom, which for N = 4 and parameter, s = 3, is ν =1 for all models except the Linear Exponential with ν =2. Also there is the value of (-L) and the closely related Akaike measure

$$\text{AIC} = -2(L - s) \qquad [12]$$

Finally some direct error measures used in regression analysis are the sums of errors squared

$$SS = \sum_{i=1}^{N} n_i(y_i - P_i)^2 \qquad [13]$$

and R_a^2, the adjusted R^2 statistic

$$R_a^2 = 1-(1-R^2)\frac{(N-1)}{(N-\nu)} \quad [14]$$

$$R^2 = 1-\frac{SS}{TSS} \quad [15]$$

$$TSS = \sum_{i=1}^{N} n_i(P_i-\overline{P_i})^2 \quad [16]$$

with TSS the total sums of squares and $\overline{P_i}$ the mean of P_i, i = 1,N.
A summary of all the measures is given in Supplemental Material, Table 2

Considering first the p-values, it is conventional in large sample sizes to expect that p should be >= 0.05. On this basis, only the Log Normal and Log Logistic would be accepted. Of the remaining four models it is to be noted that there is some inconsistency between p for X^2 and G^2 when considering models in pairs: p for Weibull is greater than that for Gamma in X^2 but reversed with p for G^2. The same can be said for the Truncated Logistic with Gamma and Weibull with Truncated Logistic.

Looking at (-L), AIC and SS values, the Log Normal and Log Logistic have lower values and probably better fits than the remaining four models which have little to distinguish between them.

The final adjusted-R_a^2 values are all consistently high suggesting good fitting for all models. As Draper and Smith (1966, p.92) point out, this is a rather gross indicator, but, as Payne et al. (1993, p.311) note, if R_a^2 were negative this would indicate a very poor fit, which is not the case here.

Because the measures considered are all designed for large sample sizes and are not necessarily reliable for very small samples, it is not straightforward to select a minimum of three good-fitting models. However, on balance, all models are considered acceptable. Therefore, based on the arithmetic average of all the $d_{0.05}$ values in Supplemental Material, Table 1, the $BMD_{0.05}$ is 0.362 mg/kg/day.

Supplemental Material, Table 1. Parameters and $d_{0.05}$ values for the suite of models.

Model	c	q_1	q_2	$d_{0.05}$
Weibull	0.047	3.3381	1.1467	0.250
Gamma	0.0157	1.9762	1.5541	0.382
Log Normal	0.0174	2.1432	0.8021	0.573
Log Logistic	0.0183	2.1541	2.3124	0.603
Truncated Normal	-	-	-	-
Truncated Logistic	0.0143	2.6048	-1.8289	0.197
Linear Exponential	0.0140	3.2753	-	0.168

Supplemental Material, Table 2. Goodness of fit measures.

Model	SS	R_a^2	X^2(p)	G^2(p)	-L	AIC
Weibull	1.093	0.9893	10.898 (0.001)	9.167 (0.003)	49.526	105.051
Gamma	0.684	0.9933	15.104 (0.0001)	7.578 (0.006)	48.730	103.461
Log Normal	0.125	0.9988	2.141 (0.143)	1.703 (0.192)	45.793	97.587
Log Logistic	0.032	0.9997	0.753 (0.385)	0.639 (0.424)	45.261	96.522
Truncated Logistic	1.222	0.9880	10.234 (0.001)	9.790 (0.002)	49.837	105.674
Linear Exponential	1.420	0.9792	8.733 (0.013)	10.097 (0.006)	49.990	103.981

REFERENCES

Draper NR, Smith H. 1966. Applied Regression Analysis, 2nd ed. New York: John Wiley.

Miller AJ. 2004. Module: Nelder-Mead simplex algorithm, minim.f90. Available: http://users.bigpond.net.au/amiller/ [accessed 2 June, 2004].

NHMRC. 1999. Toxicity Assessment for Carcinogenic Soil Contaminants. Canberra, Australia: National Health and Medical Research Council. Available: http://www.health.gov.au:80/nhmrc/publications/synopses/eh21syn.htm [accessed 9 June 2004].

Payne RW, Lane PW, Ainsley AE, Bicknell KE, Digby PGN, Harding SA, et al. 1993. Genstat 5 Release 3 Reference Manual. Oxford, UK: Clarendon Press.

Read TRC, Cressie NAC. 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data. New York: Springer Verlag.